DOCUMENT RESUME

ED 445 058 TM 031 674

AUTHOR Buckendahl, Chad W.; Impara, James C.; Plake, Barbara S.

TITLE Evaluating District Performance without a Common Assessment: Nebraska's Accountability Model.

PUB DATE 2000-06-00

NOTE 16p.; Paper presented at the Annual National Conference of the Council of Chief State School Officers on Large-Scale Assessment (30th, Snowbird, UT, June 25-28, 2000).

PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS *Academic Achievement; *Accountability; Educational Assessment; Elementary Secondary Education; *Evaluation Methods; *Performance Factors; *School Districts; State Programs; *Test Results

IDENTIFIERS *Nebraska

ABSTRACT

Many states use a statewide assessment strategy to evaluate districts on common measures. Because districts in Nebraska are not measured on common instruments, comparisons of district performance are difficult. The accountability model described in this paper contains three components related to school district performance on academic content standards: (1) student performance on district assessments; (2) technical quality of district assessments; and (3) challenge adjustments based on district demographic characteristics. The model is proposed based on a previous study that examined the appropriate component contributions of the model, the utility of creating composite scores for accountability, and the stability of district classification decisions. The current accountability model does not include a component for improvement over time and is suggested as a strategy for low stakes accountability systems. (Contains 23 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 445 058

Evaluating district performance without a common assessment: Nebraska's accountability model

Chad W. Buckendahl

James C. Impara

Barbara S. Plake

Buros Center for Testing

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

C.W. Buckendahl

1

Paper presented at the annual Large Scale Assessment Conference in Snowbird, Utah

June, 2000

TM031674

BEST COPY AVAILABLE

Abstract

Many states use a statewide assessment strategy to evaluate districts on common measures. Because districts in Nebraska are not measured on common instruments, comparisons of district performance are difficult. The accountability model described in this paper contains three components related to school district performance on academic content standards. Student performance on district assessments, technical quality of district assessments, and challenge adjustments based on district demographic characteristics comprise the three components of the Nebraska model. A study was conducted that examined the appropriate component contributions of the model; the utility of creating composite scores for accountability; and the stability of district classification decisions. The current accountability model does not include a component for improvement over time and is suggested as a strategy for low stakes accountability systems.

Evaluating district performance without a common assessment:

Nebraska's accountability model

Educational accountability is a common topic discussed among educators and administrators nationwide. It has become evident, though, that control over methods of accountability has shifted from the local jurisdiction (school districts) to the state jurisdiction (state departments of education and legislative entities). The shift is not surprising because popular media have given increased attention to educational accountability. One reason may be a general belief that public education has not lived up to the lofty expectations that have been placed on it.

Although it has received renewed attention, research on accountability systems is not new. Scholars have historically focused on outcome measures and the decisions that are generally associated with higher stakes accountability systems (Tyler, 1973; Dyer, Linn, & Patton, 1968). The validity information provided by those measures represents a critical element in evaluating district performance. Whether using norm- or criterion-referenced instruments, the alignment and proper use of those instruments relative to the desired educational objectives is essential to an equitable accountability system.

Additional problems arise when the scores are reported and schools' or districts' performance is compared. With a common assessment, states have rank ordered school districts based on test performance at individual grades and content areas (e.g., Georgia) or on a composite index of district performance that considers demographic characteristics (e.g., Kentucky) or one that does not (e.g., Texas). The rank ordering may not be meaningful, though, without also considering some of the other characteristics that may affect performance (Guskey & Kifer, 1990).

This paper describes an accountability model that was designed without a common assessment and is used to evaluate district performance on content standards in Nebraska. Components of the model include: 1) estimates of student performance on district assessments; 2) ratings of the technical quality of district assessments; and 3) challenge adjustments that give credit to districts that possess student characteristics that present high levels of challenge to district success.

The Nebraska Accountability Model

As part of the Assessment and Accountability Plan adopted by the Nebraska Board of Education, school districts will be evaluated on how well they are meeting state adopted content standards. A study was conducted that focused on the development of Composite Scale Scores (CSS) underlying the School Performance Ratings (SPR) that will be used to classify school districts on the desired evaluation criteria and to disseminate the results of the evaluation to appropriate stakeholders in a clear, understandable fashion. It was important that the components of the CSS and subsequent SPR classification system be related to the goals of the evaluation, contain components that were related to measuring student achievement, and provided all school districts an equivalent opportunity to perform well on the classification scale.

According to the Nebraska Department of Education, the SPR will initially contain three components with a fourth component added after the model has been in place for more than a year (Christensen, 1999). Each component was selected to be consistent with the adopted state assessment model and the relevance of each in measuring and explaining district performance on the state content standards. Another important consideration in the development of the CSS and SPR was a need for school districts to have equivalent opportunities to succeed within the

proposed framework. This is why a rating system, rather than a rank ordering system, was chosen. If a numerical index that ranked districts was created, differences among those rankings may not be meaningful, yet could be erroneously interpreted as such. The following sections briefly describe the three components of the Nebraska accountability model.

Component 1: Student Performance on District Assessments

Districts will provide information to an external review body describing student performance relative to the district's assessments. The specific strategies that districts use to determine the percentage of students meeting content standards will be provided in their assessment plans. Student performance described in a district assessment report (defined as the percentage of students meeting or exceeding a collective set of standards) represents the first component of the SPR.

The inclusion of student performance in a school or district evaluation model is common. Currently, forty-one states use assessment scores of some type as part of their accountability system (Mather, 1999). The assessments from which these scores are procured range from norm-referenced (Florida Department of Education, 2000) to criterion-referenced (Connecticut Department of Education, 1999) to a combination of norm- and criterion-referenced assessments (Louisiana Department of Education, 1999). In contrast to other states, Nebraska's model is district-specific, meaning that it is possible for each district to have a different combination of norm-referenced, criterion-referenced, and other assessments as part of its plan. Each of Nebraska's 550+ district plans may be unique.

Component 2: Technical Quality of District Assessments

School districts will also be required to provide documentation to an external review body that describes their overall assessment plan and contains information about the technical quality of their assessment strategies for measuring student performance on Nebraska's content standards (Nebraska Department of Education, 1999). These assessment strategies will initially measure student performance in Language Arts (reading, writing, speaking, and listening) for grades 4, 8, and 11. The review body will use a variety of indicators and criteria for assessing the technical quality and adequacy of the district's assessment plan and instruments.

The review body will evaluate the districts' assessments and rate the technical quality of the assessments into one of five categories. A technical quality rubric developed by the Buros Center for Testing specifies the psychometric characteristics necessary to achieve a given technical quality classification (Plake, 2000). The technical quality of district assessments as evaluated by this rubric represents the second component in the SPR.

The rationale for including technical quality in the SPR is that it is necessary for districts to demonstrate the psychometric soundness of the methods they are using to measure student performance. Because the Nebraska Assessment and Accountability model does not rely on a single statewide assessment, each district will be responsible, in part, for the technical quality of the assessments they are using. If districts were only asked to provide student performance estimates without some assurance that the strategies or instruments they are using to measure performance meet technical standards, it would be difficult to meaningfully interpret the results.

The technical quality criteria in the rubric represent characteristics of high quality measurement that are applicable in a school district (see for example, Traub, 1994; Anastasi,

1988; Ebel, 1979). Psychometric characteristics such as alignment to standards, opportunity to learn, bias review, developmental appropriateness, consistency in scoring, and appropriate mastery levels provide evidence of the technical quality of school districts' assessments. It follows, then, that higher technical quality assessments will produce student performance estimates that will likely be more credible than estimates produced from lower quality assessments. However, students' performance on a district's assessments and the technical quality of those assessments by themselves may not adequately represent a district's performance on the content standards. Therefore, additional factors that may affect district performance were considered in Nebraska.

Component 3: Challenge Adjustments

The technical quality of districts' assessments and student performance relative to those assessments do not consider other demographic factors that may have a negative impact on students' academic performance. Acknowledging the potential effects of these factors allows that school districts are not be able to control all factors related to its performance in the accountability model. To explain the potential impact these factors may have on student performance, adjustments will be made to a district's composite scale score. The indicators for the adjustments were selected following discussions with the Nebraska Department of Education and a review of relevant literature regarding indicators of educational performance (Mather, 1999; Reeves, 1998; Nelson, Ysseldyke, & Thurlow, 1998; Jaeger, 1979; Irvine, 1968; Cohen, 1968; Bauer, 1966).

The challenge adjustments are determined in a normative fashion, meaning that they can be used to describe the relative position of a district based on its proportion of students with specified demographic characteristics that may suppress scores on achievement tests. A majority of states that have incorporated some demographic characteristics into their accountability systems have chosen ones such as parents' level of education, dropout rate, student expenditures, student behavior, teacher/student ratio, learning or other disabilities, language barriers, mobility, and socioeconomic status (Mather, 1999). Because many of these indicators are difficult to measure, researchers use variables that may serve as proxies for each. A proxy variable is one that is believed to represent the construct to be measured. For example, teenage birth rate has been used as a proxy for the emphasis a community places on education (Reeves, 1998). The proxy variables that were selected for this model were as follows:

1. Percentage of students in free or reduced meal programs – proxy for socioeconomic status (SES).
2. Percentage of students with an Individual Education Plan (IEP) – proxy for learning or other disabilities.
3. Percent of students classified as Limited English Proficiency (LEP) or receiving English as Second Language (ESL) services – proxy for language acquisition barriers.
4. Ratio of average daily membership to enrollment – proxy for mobility.

Of the four factors selected for the Nebraska accountability model, three were common to many states (King & Mathers, 1999; Cornett & Gaines, 1997). Only mobility was not a widely used variable nationally. Mobility was operationally defined here as a district's ratio of average daily membership to enrollment. Only four states currently collect data on student mobility

(Mather, 1999). In an accountability system, the collection of student mobility information may be important when districts are evaluated on their students' performance. If students have enrolled in a district, but are not currently on the roster; they may have dropped out or left the district. Conversely, new students who have not been taught in the school or district will not have had an opportunity to learn the content on which they are tested. This potential influx of students may have an adverse affect on a district's evaluation if these students perform poorly because of the lack of control the district has over a student's learning opportunities[1].

The goal of the adjustment is to recognize and reward districts that have demographic characteristics in their student populations that represent challenges to the district's ability to meet or exceed the content standards without penalizing districts that do not have similar student populations. Only districts that are classified above the state average in any of the identified challenge areas had the opportunity to receive adjustments on the composite scale score. The challenge adjustment is intended only to benefit districts with challenging populations by making an adjustment to their CSS and potentially the subsequent SPR.

The intent is not to sanction districts that do not have a large number of challenging students. Because the scale is normative, though, some districts will benefit from an adjustment from this scale while others will not. However, because districts are classified into rating categories, not rank ordered, districts within rating categories are not compared. All districts have an equal opportunity to be classified in any rating category whether or not they receive an adjustment.

[1] It should be noted that the proxy variable for mobility does not consider within district mobility. Because the evaluation system considers district performance, a district level proxy variable was used.

Caution is urged when interpreting a scale that makes adjustments to an overall scale or rating based on uncontrollable variables. One problem when adjusting district ratings based on demographic characteristics is that such adjustments may confound interpretation of results because a district with lower test scores but a higher performance rating due to an adjustment may be classified above "satisfactory." "Students can't read any better because of the adjustment" (Irvine, 1999). Additionally, because the adjustment involves variables that districts cannot control, adjusted scores should not be rank ordered because the scores do not have the same meaning for all districts (Reeves, 1998). As stated above, the Nebraska accountability plan uses a rating, not rank ordering procedure for evaluating the performance of districts.

Future Component 4: Improvement over Time

The fourth component of the SPR will be a district's improvement over time. The improvement component is sometimes called a district's Adequate Yearly Progress (AYP). In many states it is examined as a two or three year block as opposed to a single year. This is done because individual class differences from one year to the next may account for slight variations in a district's performance. By aggregating over time, the estimates that result from this analysis are likely to be more stable. At the outset of Nebraska's accountability system this component cannot be measured and therefore a discussion of the component is not included in this paper.

Combining the Components

The general mathematical model that describes the combination of the three components is as follows:

$$(A \times B) + C_1 + C_2 + C_3 + C_4$$

In this model, "A" represents a district's student performance and "B" represents the technical quality rating of a district's assessments. Challenge adjustments for each of the demographic variables are made when districts are eligible to receive adjustments. The interaction between student performance and technical quality was selected because of the dependent relationship that exists between the quality of a district's assessments and the performance on those assessments. In a state that uses common assessments for all districts, technical quality is essentially the same. Nebraska's assessment system allows districts to create and/or select their own assessments, therefore an additional measure of quality is needed. Variations of this model were considered in a study that compared appropriate contributions of the model components, utility of creating the composite scores, and the stability of model classification decisions prior to recommending this model.

Conclusions and Implications

In a state that relies on district level assessment strategies to measure student performance on state adopted content standards, comparing districts' resultant performance is difficult. This paper described a strategy for a low-stakes state accountability system that does not rely on common statewide assessment strategies. Two additional considerations beyond traditional student performance measures are also included in the accountability model.

Given the uniqueness of districts' assessment strategies, the technical quality of a district's assessments is related to the student performance in that the assessment strategies must be sound for performance on them to be credible. This technical quality component represents a key contribution of this accountability model to practice. So, too, student performance is related to the demographic characteristics of the student population. Consideration of these

characteristics through challenge adjustments allows for a more equitable comparison of district performance in the state.

The current accountability system does not consider improvement over time in the general accountability model. This will need to be updated when at least one year has passed and another study is conducted to determine the appropriate contribution of this fourth component. Another assumption under which this model was developed was that there would be no rewards or sanctions beyond the school performance rating (SPR) and subsequent public opinion regarding that rating. By maintaining a low-stakes environment, less pressure will be placed on district administration and educators from sources outside their district to succeed on the state assessment at all costs. It is with this philosophy in mind that may allow Nebraska to avoid the pitfalls of other accountability systems that have placed greater consequences on performance.

References

Anastasi, A. (1988). *Psychological Testing*. New York, NY: Macmillan Publishing Co., Inc.

Bauer, R. (Ed.). *Social indicators*. Cambridge, MA: Massachusetts Institute of Technology Press, 1966.

Christensen, D. Personal communication. December 8, 1999.

Cohen, W.J. (1968). Social indicators: Statistics for public policy. *The American Statistician, 22*, 14.

Connecticut Bureau of Research, Evaluation, and Student Assessment (1999). *State Education Accountability Systems: Connecticut*. Washington, DC: Council of Chief State School Officers.

Cornett, L.M. & Gaines, G. (1997). *Accountability in the 1990s: Holding Schools Responsible for Student Achievement*. Atlanta, GA: Southern Regional Education Board.

Dyer, H.S., Linn, R.L, & Patton, M.J. (1968). *Methods of Measuring School System Performance*. Princeton, NJ: Educational Testing Service.

Ebel, R.L. (1979). *Essentials of educational measurement (3rd ed.)*. Englewood Cliffs, NJ: Prentice-Hall.

Florida Department of Education. (2000). *Florida Department of Education Website*. Website can be accessed at http://www.firn.edu/doe.

Guskey, T.R. & Kifer, E.W. (1990). Ranking school districts on the basis of statewide test results: Is it meaningful or misleading? *Educational Measurement: Issues and Practice, 9(1)*, 11-16.

Irvine, D.J. (1968, December). *Performance indicators in education*. Paper presented at the Commissioners Conference on State and National Assessment. Albany, NY.

Irvine, D.J. Personal communication. December 14, 1999.

Jaeger, R.M. About educational indicators: Statistics on the conditions and trends in education. In L.S. Shulman (Ed.), *Review of Research in Education 6*, pp.276-315. Itasca, IL: F.E. Peacock Publishers, Inc., 1979.

King, R.A. & Mathers, J.K. (1999). Financing schools based on performance measures. *School Business Affairs, 65(1)*, 3-9.

Louisiana Department of Education. (1999). *Louisiana Department of Education Website*. Website can be accessed at http://www.doe.state.la.us.

Mathers, J. (1999). *Accountability Systems in 50 States*. Denver, CO: Education Commission of the States.

Nelson, J.R., Ysseldyke, J.E., Thurlow, M.L. (1998). *Desired characteristics for state and school district accountability reports: Technical Report 22*. Minneapolis, MN: National Center on Educational Outcomes.

Nebraska Department of Education (1999). School-based teacher-led assessment and reporting system (STARS): A planning guide for Nebraska schools. Lincoln, NE: Nebraska Department of Education.

Plake, B.S. (2000). Technical Quality Rubric for Nebraska's School Performance Ratings. Lincoln, NE: Buros Center for Testing.

Reeves, E.B. (1998). *Do contextual effects bias Kentucky school district accountability index scores? Occasional Paper No. 1*. Morehead State University: R&D Center for Students,

Schools, and Communities.

Texas Department of Education (2000). *Texas Department of Education Website*. Website can be accessed at http://www.tea.state.tx.us/perfreport.

Traub, R. (1994). *Reliability for the social sciences: Theory and applications*. Thousand Oaks, CA: Sage Publications.

Tyler, R.W. (1973). Testing for Accountability. In R.W. Hostrop, J.A. Mecklenburger, & J.A. Wilson (Eds.), *Accountability for Educational Results*, pp. 159-162. Hamden, CT: Linnet Books.